4. Markov property continued

- Markov property of undirected graphs
- Markov property of directed graphs
- Markov property of factor graphs
- Independence maps
Markov property of undirected graphical models

- consider a distribution $\mu(x)$ that factorizes according to an undirected graphical model on $G = (V, E)$,

$$
\mu(x) = \frac{1}{Z} \prod_{c \in C} \psi_c(x_c)
$$

where $C$ is the set of all maximal cliques in $G$

- for any disjoint subsets $A, B, C \subseteq V$, $\mu(x)$ satisfy $x_A - x_B - x_C$ whenever $B$ separates $A$ and $C$
**example:** color the nodes with \( \{ R, G, B \} \) such that no adjacent node has the same color

\[
\mu(x) = \frac{1}{Z} \prod_{(i,j) \in E} \mathbb{I}(x_i \neq x_j)
\]

for a node \( i \), if we condition on the color of the neighbors of \( i \), color of \( i \) is independent of the rest of the graph

Markov property
Hammersley-Clifford theorem

- (pairwise) if positive \( \mu(x) \) satisfies all conditional independences implied by a graph \( G \) without any triangles, then we can find a factorization

\[
\mu(x) = \frac{1}{Z} \prod_{(i,j) \in E} \psi_{ij}(x_i, x_j)
\]

- (general) if positive \( \mu(x) \) satisfies all conditional independences implied by a graph \( G \), then we can find a factorization

\[
\mu(x) = \frac{1}{Z} \prod_{c \in C} \psi_{c}(x_c)
\]

- there are conditional independencies that cannot be represented by an undirected graphical model, but is represented by a Bayesian network

Example: \( \mu(x) = \mu(x_1)\mu(x_3)\mu(x_2|x_1, x_3) \)

\[x_1 \text{ and } x_3 \text{ are independent} \quad \text{ no independence}\]
Markov property of directed graphical models

Examples

1. \[ \mu(x) = \mu(x_1)\mu(x_2|x_1)\mu(x_3|x_2) \]
   
   since \( \mu(x_3|x_2) = \mu(x_3|x_1, x_2) \), we have \( x_1 \!-\! x_2 \!-\! x_3 \)

2. \[ \mu(x) = \mu(x_2)\mu(x_1|x_2)\mu(x_3|x_2) \]
   
   since \( \mu(x_3|x_2) = \mu(x_3|x_1, x_2) \), we have \( x_1 \!-\! x_2 \!-\! x_3 \)

3. \[ \mu(x) = \mu(x_1)\mu(x_3)\mu(x_2|x_1, x_3) \]

   \( x_1 \) and \( x_3 \) are independent but not \( x_1 \!-\! x_2 \!-\! x_3 \)

4. \[ \mu(x) = \mu(x_1)\mu(x_3)\mu(x_2|x_3) \]
for any distribution $\mu(x)$ that factorizes according to a directed acyclic graph (DAG) $G$, and for any disjoint subsets $A, B, C \subseteq V$, we can test whether $x_A - x_B - x_C$ using **Bayes ball algorithm**:

1. shade all nodes in $B$
2. place a ball at each node in $A$
3. let balls bounce around in the graph following the rules shown on the right (Remark: balls do not interact)
4. if no ball can reach $C$, then $x_A - x_B - x_C$

\[ A = \{4\} \]
\[ B = \{3, 5\} \]
\[ C = \{6\} \]
for any distribution $\mu(x)$ that factorizes according to a directed acyclic graph (DAG) $G$, and for any disjoint subsets $A, B, C \subseteq V$, we can test whether $x_A - x_B - x_C$ using **Bayes ball algorithm**:

1. shade all nodes in $B$
2. place a ball at each node in $A$
3. let balls bounce around in the graph following the rules shown on the right (Remark: balls do not interact)
4. if no ball can reach $C$, then $x_A - x_B - x_C$

\[ A = \{4\} \]
\[ B = \{3, 5\} \]
\[ C = \{6\} \]
for any distribution $\mu(x)$ that factorizes according to a directed acyclic graph (DAG) $G$, and for any disjoint subsets $A, B, C \subseteq V$, we can test whether $x_A - x_B - x_C$ using **Bayes ball algorithm**:

1. shade all nodes in $B$
2. place a ball at each node in $A$
3. let balls bounce around in the graph following the rules shown on the right (Remark: balls do not interact)
4. if no ball can reach $C$, then $x_A - x_B - x_C$

**Markov property**

$A = \{4\}$
$B = \{3, 5\}$
$C = \{6\}$
for any distribution $\mu(x)$ that factorizes according to a directed acyclic graph (DAG) $G$, and for any disjoint subsets $A, B, C \subseteq V$, we can test whether $x_A \rightarrow x_B \rightarrow x_C$ using **Bayes ball algorithm**:

1. shade all nodes in $B$
2. place a ball at each node in $A$
3. let balls bounce around in the graph following the rules shown on the right (Remark: balls do not interact)
4. if no ball can reach $C$, then $x_A \rightarrow x_B \rightarrow x_C$ is blocked

**Diagram:**

![Bayes ball algorithm diagram](image)

**Example:**

$A = \{4\}$

$B = \{3, 5\}$

$C = \{6\}$
for any distribution \( \mu(x) \) that factorizes according to a directed acyclic graph (DAG) \( G \), and for any disjoint subsets \( A, B, C \subseteq V \), we can test whether \( x_A \rightarrow x_B \rightarrow x_C \) using **Bayes ball algorithm**:

1. shade all nodes in \( B \)
2. place a ball at each node in \( A \)
3. let balls bounce around in the graph following the rules shown on the right (Remark: balls do not interact)
4. if no ball can reach \( C \), then \( x_A \rightarrow x_B \rightarrow x_C \) blocked

\[
\begin{array}{c|c}
\text{1} & \text{2} \\
\text{3} & \text{4} \\
\text{5} & \text{6} \\
\text{7} & \\
\end{array}
\]

\( A = \{4\} \)
\( B = \{3, 5\} \)
\( C = \{6\} \)
for any distribution $\mu(x)$ that factorizes according to a directed acyclic graph (DAG) $G$, and for any disjoint subsets $A, B, C \subseteq V$, we can test whether $x_A \!-\! x_B \!-\! x_C$ using **Bayes ball algorithm**:

1. shade all nodes in $B$
2. place a ball at each node in $A$
3. let balls bounce around in the graph following the rules shown on the right (Remark: balls do not interact)
4. if no ball can reach $C$, then $x_A \!-\! x_B \!-\! x_C$ blocked

\[ A = \{4\} \]
\[ B = \{3, 5\} \]
\[ C = \{6\} \]
for any distribution $\mu(x)$ that factorizes according to a directed acyclic graph (DAG) $G$, and for any disjoint subsets $A, B, C \subseteq V$, we can test whether $x_A - x_B - x_C$ using **Bayes ball algorithm**:

1. shade all nodes in $B$
2. place a ball at each node in $A$
3. let balls bounce around in the graph following the rules shown on the right (Remark: balls do not interact)
4. if no ball can reach $C$, then $x_A - x_B - x_C$

**Diagram:**

- $A = \{4\}$
- $B = \{3, 5\}$
- $C = \{6\}$

Markov property
for any distribution $\mu(x)$ that factorizes according to a directed acyclic graph (DAG) $G$, and for any disjoint subsets $A, B, C \subseteq V$, we can test whether $x_A \!-\! x_B \!-\! x_C$ using **Bayes ball algorithm**:

1. shade all nodes in $B$
2. place a ball at each node in $A$
3. let balls bounce around in the graph following the rules shown on the right (Remark: balls do not interact)
4. if no ball can reach $C$, then $x_A \!-\! x_B \!-\! x_C$

---

**Diagram:**

- **Nodes:** 1, 2, 3, 4, 5, 6, 7
- **Edges:**
  - 1 -> 2, 2 -> 1, 2 -> 3, 3 -> 1, 3 -> 5, 5 -> 3, 5 -> 6, 6 -> 5, 6 -> 7, 7 -> 6, 4 -> 5, 5 -> 4, 4 -> 7, 7 -> 4
- **Shaded Nodes:** 3, 5
- **Balls:** 1, 2, 4, 5

**Sets:**

- $A = \{4\}$
- $B = \{3, 5\}$
- $C = \{6\}$
for any distribution $\mu(x)$ that factorizes according to a directed acyclic graph (DAG) $G$, and for any disjoint subsets $A, B, C \subseteq V$, we can test whether $x_A - x_B - x_C$ using Bayes ball algorithm:

1. shade all nodes in $B$
2. place a ball at each node in $A$
3. let balls bounce around in the graph following the rules shown on the right (Remark: balls do not interact)
4. if no ball can reach $C$, then $x_A - x_B - x_C$

\[A = \{4\}\]
\[B = \{3, 5\}\]
\[C = \{6\}\]
for any distribution $\mu(x)$ that factorizes according to a directed acyclic graph (DAG) $G$, and for any disjoint subsets $A, B, C \subseteq V$, we can test whether $x_A - x_B - x_C$ using **Bayes ball algorithm**:

1. shade all nodes in $B$
2. place a ball at each node in $A$
3. let balls bounce around in the graph following the rules shown on the right (Remark: balls do not interact)
4. if no ball can reach $C$, then $x_A - x_B - x_C$ blocked

Q. do we need bouncing?

Q. when do we stop?

Markov property

$A = \{4\}$

$B = \{3, 5\}$

$C = \{6\}$

implies $x_4 - \{x_3, x_5\} - x_6$
if a distribution $\mu(x)$ factorizes according to a DAG $G$, i.e.

$$\mu(x) = \prod_{i \in V} \mu_i(x_i | x_{\pi(i)})$$

then $\mu(x)$ satisfies all the conditional independencies obtainable by Bayes ball

if a distribution $\mu(x)$ satisfies all the conditional independencies obtainable by Bayes ball on a a DAG $G$, then we can find a factorization of $\mu(x)$ on $G$

there are conditional independencies that cannot be represented by a Bayesian network, but is represented by an undirected graphical model,

**Example:** $\mu(x) = \psi_{12}(x_1, x_2)\psi_{23}(x_2, x_3)\psi_{34}(x_3, x_4)\psi_{41}(x_4, x_1)$

$\begin{align*}
 x_1 - \{x_2, x_4\} - x_3, & \quad x_2 - \{x_1, x_3\} - x_4
\end{align*}$
Markov property of factor graphs

Factor graphs are more ‘fine grained’ than undirected graphical models

\[ ψ(x_1, x_2, x_3) \quad ψ_{12}(x_1, x_2)ψ_{23}(x_2, x_3)ψ_{31}(x_3, x_1) \quad ψ_{123}(x_1, x_2, x_3) \]

all three encodes same independencies, but different factorizations
(in particular $3|\mathcal{X}|^2$ vs. $|\mathcal{X}|^3$)

- set of independencies represented by MRF is the same as FG
- but FG can represent a larger set of factorizations
- for a factor graph $G = (V, F, E)$, for any disjoint subsets $A, B, C \subseteq V$, $μ(x)$ satisfy $x_A−x_B−x_C$ whenever $B$ separates $A$ and $C$
Independence maps (I-maps)

which graphical model is ‘good’?

- let $\mathcal{I}(G)$ denote all conditional independencies implied by a graph $G$
- let $\mathcal{I}(\mu)$ denote all conditional independencies of a distribution $\mu(\cdot)$
- $G$ is an **I-map** of $\mu$ if $G$ captures (some) independencies of distribution $\mu(x)$

$$\mathcal{I}(G) \subseteq \mathcal{I}(\mu)$$

- given $\mu(x)$, can we construct a $G$ that captures as many independencies as possible?
- $G$ is a **minimal I-map** for $\mu(x)$ if
  - $G$ is an I-map for $\mu(x)$, and
  - removing a single edge from $G$ causes the graph to no longer be an I-map
Constructing minimal I-maps

Constructing minimal I-maps for Bayesian network

1. choose an (arbitrary) ordering of $x_1, x_2, \ldots, x_n$
2. consider a directed graph for Bayes rule

$$
\mu(x) = \mu(x_1)\mu(x_2|x_1)\cdots\mu(x_n|x_1, \ldots, x_{n-1})
$$

3. for each $i$, select its parents $\pi(i)$ to be the minimal subset of $
\{x_1, \ldots, x_{i-1}\}$ such that

$$
x_i - x_{\pi(i)} - \{x_1 \ldots, x_{i-1}\} \setminus x_{\pi(i)}
$$

• minimal I-map for BN is not unique

Constructing minimal I-maps for Markov random fields

1. consider a complete undirected graph
2. for each edge $(i, j)$, remove the edge if $i$ and $j$ are conditionally independent given all the rest of nodes

• if $\mu(x) > 0$, then the resulting graph is the unique minimal I-map of $\mu$

Markov property
Perfect maps (P-maps)

$G$ is a **perfect map (P-map)** for $\mu(x)$ if

$$\mathcal{I}(G) = \mathcal{I}(\mu)$$

set of all distributions on $V$

examples:

trees have efficient inference algorithms, which plays a crucial role in developing efficient algorithms for other graphs as well
Trees

- **undirected tree** is a connected undirected graph with no cycle
- **directed tree** is a connected directed graph where each node has at most one parent
- conversion from MRF on tree to BN on tree: take any node as root and start ‘directing’ edges away from the root
  - non-unique
  - all conversions result in the same set of independencies
- **example**: Markov chain, hidden Markov models
- exact inference is extremely efficient on trees

Markov property
Chordal graphs

- consider a graph $G = (V, E)$ which may be directed or undirected
- we say $(i_1, \ldots, i_k)$ from a trail in the graph $G = (V, E)$ if for every $j \in \{1, \ldots, k - 1\}$, we have either $(i_j, i_{j+1}) \in E$ or $(i_{j+1}, i_j) \in E$ (direction does not matter)
- a loop is a trail $(i_1, \ldots, i_k)$ where $i_k = i_1$
- a graph (directed or undirected) is chordal if the longest minimal loop is a triangle

when do we lose nothing in converting an undirected graph to a directed one?

- for an undirected graph $G$ and a directed graph $D$
- if $\mathcal{I}(G) = \mathcal{I}(D)$ then $G$ is a chordal graph
- if $G$ is a chordal graph then exists a $D$ such that $\mathcal{I}(G) = \mathcal{I}(D)$
- use triangulation to construct such a directed graph $D$ from an undirected chordal graph $G$, which we will learn later in this course
Moralization

- one way to convert a directed graph $D$ to an undirected graph $G$ is \textbf{moralization}
- resulting $G$ is an minimal \text{I-map} of $D$

1. retain all edges in $D$ and make them undirected
2. connect every pair of parents with an edge

when do we lose nothing in converting a directed graph to an undirected one?

- if $\mathcal{I}(D) = \mathcal{I}(G)$ then moralization of $D$ does not add any edges
- if moralization of $D$ does not add any edges then there exists an undirected graph $G$ such that $\mathcal{I}(D) = \mathcal{I}(G)$